

Rigorous bounds on the stationary distributions of the chemical master equation via mathematical programming

Juan Kuntz,^{1, 2, a)} Philipp Thomas,^{2, b)} Guy-Bart Stan,^{1, c)} and Mauricio Barahona^{2, d)}

¹⁾*Department of Bioengineering, Imperial College London, London, United Kingdom*

²⁾*Department of Mathematics, Imperial College London, London, United Kingdom*

(Dated: 20 February 2017)

The stochastic dynamics of networks of biochemical reactions in living cells are typically modelled using chemical master equations (CMEs). The stationary distributions of CMEs are seldom solvable analytically, and few methods exist that yield numerical estimates with computable error bounds. Here, we present two such methods based on mathematical programming techniques. First, we use semidefinite programming to obtain increasingly tighter upper and lower bounds on the moments of the stationary distribution for networks with rational propensities. Second, we employ linear programming to compute convergent upper and lower bounds on the stationary distributions themselves. The bounds obtained provide a computational test for the uniqueness of the stationary distribution. In the unique case, the bounds collectively form an approximation of the stationary distribution accompanied with a computable ℓ^1 -error bound. In the non-unique case, we explain how to adapt our approach so that it yields approximations of the ergodic distributions, also accompanied with computable error bounds. We illustrate our methodology through two biological examples: Schlögl's model and a toggle switch model.

I. INTRODUCTION

Cell-to-cell variability is pervasive in cell biology and biotechnology. This stochasticity stems from the fact that biochemical reactions inside living cells often involve biomolecules present in only few copies per cells^{1–3}. Under well-mixed conditions the chemical master equation (CME) is used to model the stochastic dynamics of interacting biochemical components in the cell. Prominent examples include gene regulatory and signalling networks, some of which are involved in cellular adaptation and cell fate decisions^{4–6}. As increasingly accurate experimental techniques become available, it is of paramount importance to develop accurate and reliable solution methods to stochastic models. By enabling the reliable inference of underlying model parameters^{7,8}, these methodologies can facilitate the identification of molecular mechanisms⁹ and the design of synthetic circuits in living cells^{10,11}.

Significant efforts have been made to investigate the stationary distributions of CMEs because they determine the long time behaviour of the underlying continuous-time Markov chain¹². While exact Monte-Carlo methods have been developed to sample from stationary distributions¹³, analytical solutions, however, are known only in a few special cases. These include single species one-step processes¹⁴, certain classes of unimolecular reaction networks^{15,16}, networks obeying the conditions of detailed balance^{17,18}, deficiency zero networks^{19,20}, and small model systems^{21,22}. Despite these efforts, the CME is generally considered intractable because, aside of sys-

tems with finite state space, it consists of infinitely many coupled equations.

One approach to circumvent this problem is to compute moments of the stochastic process. However, this approach circumvents the problem only for networks of unimolecular reactions. In all other cases the equations for the first few moments involve unknown higher moments and thus these equations also constitute an infinite number of coupled equations that cannot be solved analytically. Moment approximations, some of which require assumptions about the apriori unknown distribution solution, are commonly employed to overcome this issue^{23–26}. Few of these methods, however, admit quantified error estimates on its solution^{27,28}. Independently, mathematical programming has been employed to compute bounds on the moments of Markov processes in various contexts. All of these consider a finite set of moment equations supplemented by moment inequalities whose solutions can be bounded in terms of a linear program (LP)^{29,30} or, under stronger conditions, using a semidefinite program (SDP)^{31,32}.

Another approach is to approximate the distribution solution, either by expanding the CME using a finite number of basis functions^{33,34} or by projecting it onto finite state spaces^{35,36}. The finite state projection algorithm³⁵, for instance, solves the CME on a finite subset with an absorbing boundary. A particular advantage of this approach over other methods is that the accuracy of its solution is controlled. Specifically, the solution of the finite state projection algorithm is a lower bound on the time-dependent solution of the CME, while the probability mass in the absorbing state yields a bound on the total error of its solution. However, at long times, all of the probability mass becomes trapped into the absorbing state, and for this reason the method is not well-suited to study the CME's long-term behaviour and its stationary distributions.

^{a)}Electronic mail: jk208@ic.ac.uk

^{b)}Electronic mail: p.thomas@ic.ac.uk

^{c)}Corresponding author: g.stan@ic.ac.uk

^{d)}Corresponding author: m.barahona@ic.ac.uk

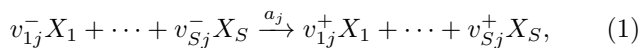
In this paper we present two mathematical programming approaches, one that yields upper and lower bounds on the moments of any stationary distribution, the other that provides bounds on the probability that the stationary distribution gives to a given subset of the state space. We build on previous methodologies developed to obtain moment bounds of polynomial diffusions using semidefinite programming³² and adapt these to discrete reaction networks with polynomial or rational propensities. In particular, we consider the first few moment equations, which are generally not closed (that is, underdetermined), and constrain their solution space using positive semidefinite inequalities that are satisfied by the moments of any admissible probability distribution.

The second approach yields convergent upper and lower bounds on the probability that the stationary distributions give to a given subset of the state space by solving LPs. By repeatedly applying the approach, we produce bounds on the complete stationary distributions and their marginals. The method considers finite but underdetermined systems of stationary equations supplemented with inequalities involving a moment bound computed using the first approach. Additionally, we quantify the error in our approach, by providing a computable bound on the total variation (or, equivalently, the ℓ^1) distance between the vector of bounds and the exact distribution solution (Lemma 3).

The paper is organised as follows. In Sec. II we review the basic equations for stationary moments and distributions. We then discuss the problem of bounding the stationary moments in Sec. III. We introduce our approach by demonstrating how analytical bounds for the first two moments can be obtained for a single-species example and then develop its generalisation to multi-species networks with rational propensities. In Sec. IV, we show how to bound event probabilities, Theorem 2. We first demonstrate the approach for single-species one-step processes for which distribution bounds can be obtained analytically. Then, we explain how to generalise the approach to bound event probabilities, distributions and marginals of arbitrary networks. In Sec. V, we exemplify the usefulness of our methods by computing the stationary distribution and marginals of a toggle switch. We conclude with a discussion of our results in Sec. VI.

II. STATIONARY MOMENTS AND DISTRIBUTIONS

We consider a generic reaction network involving S species X_1, X_2, \dots, X_S that interact via R reactions of the form



where j is the reaction index running from 1 to R , $a_j(n)$ is the propensity of the j^{th} reaction, and v_{ij}^\pm are the stoichiometric coefficients. Denoting the number of molecules at time t with $x(t) := (x_1(t), \dots, x_S(t))$, the

probability $p_n(t) = \Pr(x(t) = n)$ of observing the process in the state $n := (n_1, \dots, n_S)$ at time t satisfies the CME^{37,38}

$$\frac{dp}{dt} = Qp,$$

with initial conditions $p_n(0) = \Pr(x(0) = n)$ for each n in \mathbb{N}^S . The rate matrix Q is defined as

$$Q_{nm} := \begin{cases} a_j(m) & \text{if } m = n + v_j \\ -\sum_{j=1}^R a_j(m) & \text{if } m = n \\ 0 & \text{otherwise} \end{cases},$$

for all vectors $n \in \mathbb{N}^S$ and $v_j := (v_{1j}^+ - v_{1j}^-, \dots, v_{Sj}^+ - v_{Sj}^-)$ is the stoichiometric vector that denotes the net-change of molecule numbers incurred in the j^{th} reaction.

A stationary distribution p of the network is a distribution on \mathbb{N}^S such that if the x has law p at time zero, then x is a stationary process. Consequently, p is a fixed point of the CME, e.g. see Theorem 4.3 in Chapter 5.4 of Ref.³⁸,

$$Qp = 0. \quad (2)$$

Since the network may have more than one stationary distribution, we denote the set of solutions by \mathcal{P} . Assuming that the network is positive recurrent, the stationary distributions determine the long term behaviour of the chain¹². Verifying whether or not a network is positive recurrent consists of finding an appropriate Lyapunov function³⁹, which in certain cases can be achieved using mathematical programming⁴⁰.

Alternatively, we can consider expectations: from Eq. (2), it follows that, if a real-valued function $f(n)$ on \mathbb{N}^S satisfies some integrability conditions (for example, Condition (B1) in Appendix B), then

$$\langle Q^T f \rangle_p = 0, \quad (3)$$

where we are using vector notation $(Q^T f)(n) := \sum_{m \in \mathbb{N}^S} Q_{mn} f(m)$, see the proof of Theorem 1 in Appendix B. For networks with polynomial propensities $a_j(n)$, the equations for the moments up to moment order d are obtained by letting $f(n) := n_1^{\alpha_1} n_2^{\alpha_2} \dots n_S^{\alpha_S}$ in Eq. (3) for all $\alpha_1 + \alpha_2 + \dots + \alpha_S \leq d$. However, these equations are closed only for networks with propensities that depend at most linearly on the molecular states n , such as networks composed solely of zero-order and unimolecular reactions with mass-action propensities. For networks composed bimolecular or higher-order reactions, the equations of order d depend on moments of order higher than d , and hence these equations are not closed, that is, they form a system of underdetermined linear equation.

III. BOUNDING THE STATIONARY MOMENTS

We here develop rigorous bounds on the stationary moments of the CME. We introduce our approach for

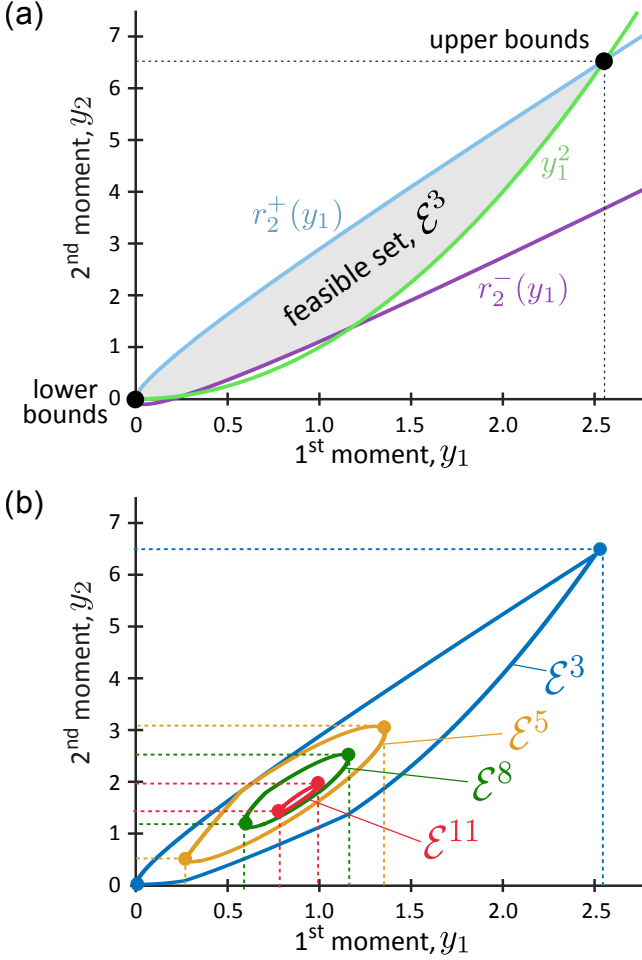


FIG. 1. **Constructing the set of possible moment values for Schlögl's model Eqs. (4)-(5).** Parameter values are $k_1 = 1$, $k_2 = 1$, $k_3 = 4/5$, and $k_4 = 1$. **(a)** The set of feasible moment values is constructed using the first moment Eq. (6) involving the first three moments, y_1 , y_2 and y_3 , whose values are constrained by positive semidefinite inequalities (see main text for details, Eqs. (7)-(8)). The area between the blue and purple lines denotes the solutions obtained from combining the moment equation with $y_1 y_3 - y_2^2 \geq 0$, the area above the green line is $y_2 - y_1^2 \geq 0$. The intersection of these sets is the feasible set of moment values \mathcal{E}^3 (grey area) given by Eq. (9). Dots show upper and lower bounds on the first and second moments (black dots). **(b)** Increasing the number of moment equations and inequalities narrows down the feasible set of moment values (areas contained in coloured outlines). The corresponding lower and upper bounds on the moments (dots) approach each other.

a model system and obtain analytically bounds on the first two moments, and then present the general method that relies on solving SDPs.

A. Moment bounds for Schlögl's model

We consider an autocatalytic network proposed by Schlögl⁴¹, that models a chemical phase transition involving a single species X ,



The propensities follow mass-action kinetics and are given by

$$\begin{aligned} a_1(n) &= k_1 n(n-1), & a_2(n) &= k_2 n(n-1)(n-2), \\ a_3(n) &= k_3, & a_4(n) &= k_4 n. \end{aligned} \quad (5)$$

If the reactions are modelled stochastically, the network has a unique stationary distribution p with either uni- or bimodal distributions depending on the parameters⁴², and all of its moments are finite for any set of positive reaction rate constants, see Appendix A.

At stationarity, the first moment equation reads

$$0 = b_1 y_0 - b_2 y_1 + b_3 y_2 - b_4 y_3, \quad (6)$$

where $y_0 := \langle 1 \rangle_p = 1$, $y_1 := \langle n \rangle_p$, $y_2 := \langle n^2 \rangle_p$, and $y_3 := \langle n^3 \rangle_p$ denote the moments of p , and $b_1 := k_3$, $b_2 := k_1 + 2k_2 + k_4$, $b_3 := k_1 + 3k_2$, $b_4 := k_2$. Instead of applying a moment closure, which assumes a particular form of the distribution, we will consider inequalities satisfied by the moments of any possible distribution.

Specifically, for any distribution \mathbb{N} with moments y_0, y_1, y_2 , and y_3 , it holds that the following two matrices are positive semidefinite⁴³ (in short $\succeq 0$)

$$\begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} \succeq 0, \quad \begin{bmatrix} y_1 & y_2 \\ y_2 & y_3 \end{bmatrix} \succeq 0. \quad (7)$$

By Sylvester's criterion the above is equivalent to

$$\begin{aligned} y_0 &\geq 0, & y_1 &\geq 0, & y_2 &\geq 0, & y_3 &\geq 0, \\ y_0 y_2 - y_1^2 &\geq 0, & y_1 y_3 - y_2^2 &\geq 0. \end{aligned} \quad (8)$$

The first five inequalities state that the moments are non-negative and that so is the distribution's variance. The last inequality gives a condition involving moments 1-3.

The set \mathcal{E}^3 of vectors $y := (y_0, y_1, y_2, y_3) \in \mathbb{R}^4$ that satisfy Eq. (6), the inequalities (8), and the normalising condition $y_0 = 1$, constrains the values the moments of p can take in the sense that $(1, \langle n \rangle_p, \langle n^2 \rangle_p, \langle n^3 \rangle_p) \in \mathcal{E}^3$. To give an explicit characterisation of this set, we restrict ourselves to the case $b_3 \geq 2\sqrt{b_2 b_4}$ (the other case can be treated similarly). A vector y satisfies the last inequality, $y_0 = 1$, and the moment equation (6) if and only if

$$\begin{aligned} 0 &\leq b_1 y_1 + b_3 y_1 y_2 - b_4 y_2^2 - b_2 y_1^2 \\ &= b_1 y_1 + (b_3 - 2\sqrt{b_2 b_4}) y_1 y_2 - (\sqrt{b_4} y_2 - \sqrt{b_2} y_1)^2. \end{aligned}$$

The set of pairs of nonnegative numbers that satisfy the above inequality and $y_2 \geq y_1^2$ can be reduced to

$$y_1 \geq 0, \quad \max\{y_1^2, r_2^-(y_1)\} \leq y_2 \leq r_2^+(y_1),$$

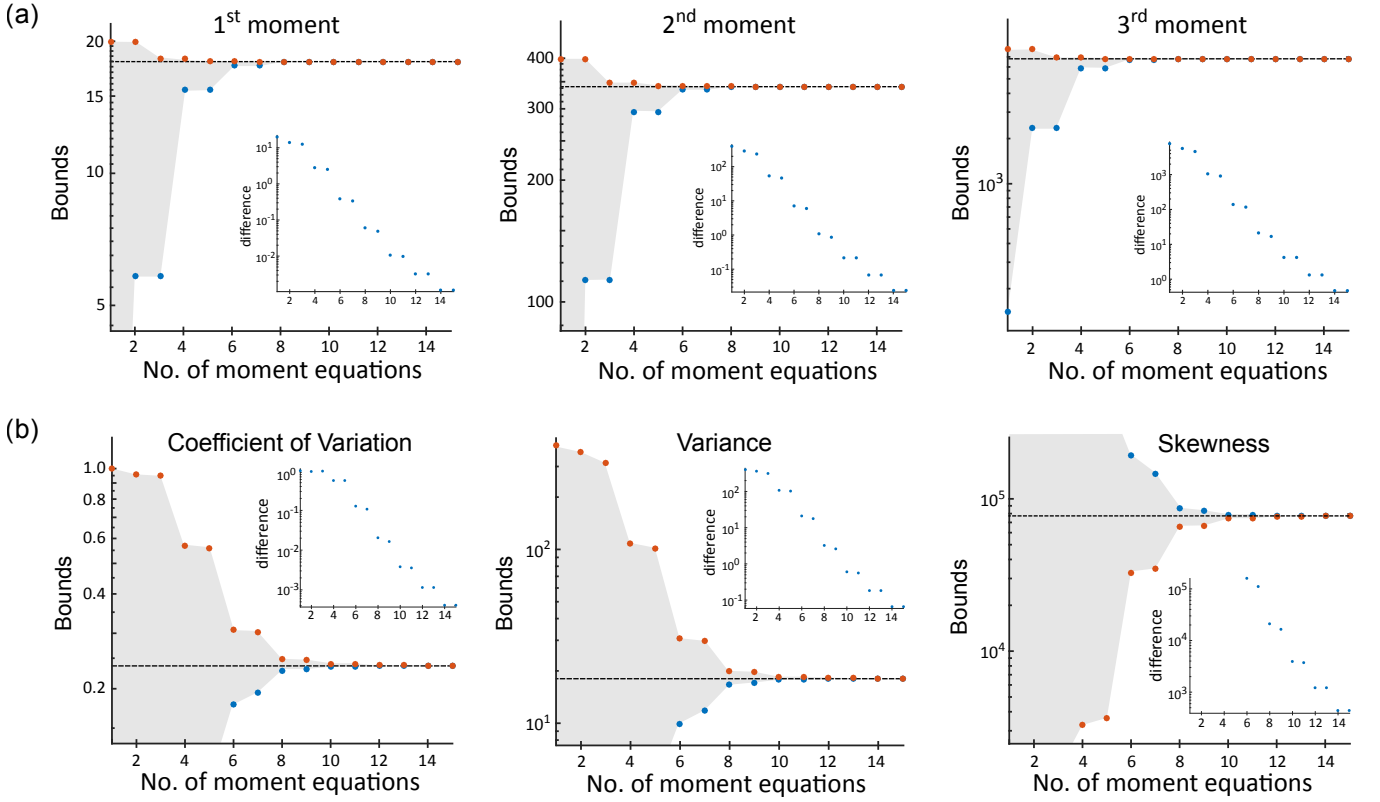


FIG. 2. **The moment bounds for Schlögl's model Eqs. (4)-(5) converge to the true moments.** (a) Upper (red) and lower (blue) bounds on the first three moments computed increasing the number of moment equations and moment inequalities (see Sec.III B 2). Inset shows that the difference between upper and lower bounds decreases with the number of moment equations used, which indicates that either bound converges to the true moment. Dashed lines indicate the limiting value approximating the true moment. (b) Using the moment bounds in (a) we compute estimates of several summary statistics such as the coefficient of variation, the variance of fluctuations and their skewness. The inset shows the difference of lower and upper bounds on these summary statistics, which indicates convergence to the true statistic with increasing number of moment equations and inequalities. Parameters are $k_1 = 6$, $k_2 = 1/3$, $k_3 = 50$, and $k_4 = 3$ leading to a unimodal distribution.

with

$$r_2^\pm(x) := \frac{b_3 y_1 \pm \sqrt{4b_1 b_4 y_1 + (b_3^2 - 4b_2 b_4) y_1^2}}{2b_4}.$$

In summary, the feasible set of moment values is given by

$$\mathcal{E}^3 = \left\{ y \in \mathbb{R}^4 : \begin{array}{l} y_0 = 1, \quad 0 \leq y_1 \leq r_4, \\ \max\{y_1^2, r_2^-(y_1)\} \leq y_2 \leq r_2^+(y_1), \\ y_3 = (b_1 - b_2 y_1 + b_3 y_2)/b_4. \end{array} \right\}. \quad (9)$$

The projection of this set onto the y_1 - y_2 plane is shown in Fig. 1a.

Because the moments of p are contained in \mathcal{E}^3 , which is clearly bounded, we have that

$$\min\{y_\alpha : y \in \mathcal{E}^3\} \leq \langle n^\alpha \rangle_p \leq \max\{y_\alpha : y \in \mathcal{E}^3\}$$

are respective upper and lower bounds on $\langle n^\alpha \rangle_p$ for $\alpha = 1, 2, 3$. To determine these bounds, we note that the functions $x \mapsto x^2$ and $x \mapsto r_2^\pm(x)$ are monotonically

increasing, such that the maximum values of y_1 and y_2 are achieved at the furthestmost northwestern intersection point of the curves $y_2 = y_1^2$ and $y_2 = r_2^+(y_1)$ (Fig. 1a dots). In other words, there is a unique maximum (y_1^*, y_2^*) with $y_2^* = r_2^+(y_1^*)$ and y_1^* is the rightmost root r_4 of the quartic polynomial $x \mapsto x(b_1 - b_2 x + b_3 x^2 - b_4 x^3)$. It hence follows that

$$0 \leq \langle n \rangle_p \leq r_4, \quad 0 \leq \langle n^2 \rangle_p \leq r_2^+(r_4),$$

which provides us with lower (albeit uninformative) and upper bounds on the first and second moments.

B. Moment bounds for multi-species reaction networks

We generalise the above approach to obtain moment bounds of multi-species networks of the form (1). Since in many applications reaction networks with rational propensities are prevalent, it is desirable to consider moment equations that allow for this type of propensity. To

this end, we introduce rational moments of the form

$$\left\langle \frac{n^\alpha}{q(n)} \right\rangle_p, \quad (10)$$

with polynomial denominator $q(n)$ (with $q(n) > 0$ for all $n \in \mathbb{N}^S$). Here, we employ multi-index notation in which multi-indices are denoted by Greek letters $\alpha := (\alpha_1, \alpha_2, \dots, \alpha_S)$, monomials are written as $n^\alpha := n_1^{\alpha_1} n_2^{\alpha_2} \dots n_S^{\alpha_S}$, and $|\alpha| := \alpha_1 + \alpha_2 + \dots + \alpha_S$. More generally, we denote by the rational moments of order d those for which $|\alpha| \leq d$ in Eq. (10).

The key observation is that by letting $f(n) = n^\alpha$ in Eq. (3), one obtains linear equations involving the rational moments with appropriately chosen denominator $q(n)$. Moreover, the power moments can be expressed as linear combinations of these moments

$$\langle n^\alpha \rangle_p = \sum_{\beta \geq 0} q_\beta \left\langle \frac{n^{\beta+\alpha}}{q(n)} \right\rangle_p,$$

where the q_β denote the polynomial coefficients such that $q(n) = \sum_{\beta \geq 0} q_\beta n^\beta$. As we show in the following

$$\inf\{f^T y : y \in \mathcal{E}_q^d\} \leq \langle n^\alpha \rangle_p \leq \sup\{f^T y : y \in \mathcal{E}_q^d\}, \quad (11)$$

where the vector f represents the polynomial coefficients of $q(n)n^\alpha$ and \mathcal{E}_q^d is the set of admissible values for the rational moments of order up to order d with denominator $q(n)$, which satisfy both the rational moment equations and certain positive semidefinite inequalities. Since the set is \mathcal{E}_q^d is defined by linear equalities and semidefinite inequalities, both the left and right hand side of (11) are SDPs, a tractable class of convex optimisation problems that can be efficiently solved computationally using standard solvers. The details of this approach are developed in the following.

1. Rational moment equations

To derive the moment equations for networks with rational propensities, we rewrite the propensities as $a_j(n) = \tilde{a}_j(n)/q(n)$, where $\tilde{a}_j(n)$ and $q(n) > 0$ are polynomials of degrees $d_{\tilde{a}_j}$ and d_q . For example, for $a_j(n) := s_j(n)/t_j(n)$, where $s_j(n)$ (resp. $t_j(n)$) are nonnegative (resp. positive) polynomials, a convenient choice is the common denominator $q(n) := \prod_{j=1}^R t_j(n) > 0$.

Choosing now $f(n) = n^\alpha$ in Eq. (3) and rearranging, we find that the rational moments satisfy the linear equations

$$\sum_{|\beta| < d_a + |\alpha|} C_{\alpha\beta} \left\langle \frac{n^\beta}{q(n)} \right\rangle_p = 0,$$

where

$$C_{\alpha\beta} := \sum_{j=1}^R \sum_{\delta \leq \beta, \delta < \alpha} \binom{\alpha}{\delta} v_j^{\alpha-\delta} \tilde{a}_{j,\beta-\delta},$$

and $\tilde{a}_{j,\beta}$ are the coefficients in the polynomials $\tilde{a}_j(n) = \sum_{\beta \geq 0} \tilde{a}_{j,\beta} n^\beta$ and $d_a := \max_j \{d_{\tilde{a}_j}\}$ is its maximum degree. Note that $\tilde{a}_{j,\alpha} = 0$ for $|\alpha| > d_{\tilde{a}_j}$ and that the second sum is taken over all δ such that $\delta_i \leq \min\{\alpha_i, \beta_i\}$ for all i and such that there is at least one i for which $\delta_i < \alpha_i$. It is clear that for polynomial propensities, i.e. $q(n) = 1$, we recover the equations for the power moments.

The equations obtained by plugging $f(n) := n^\alpha$ into Eq. (3) for all $|\alpha| \leq d - d_a + 1$ involve the rational moments of order d . We can write these equations compactly in vector notation as

$$C^d y = 0,$$

where $y_\alpha := \langle n^\alpha / q(n) \rangle_p$ for all $|\alpha| \leq d$, and $C_{\alpha\beta}^d := C_{\alpha\beta}$ for all $|\beta| \leq d$. For instance, in the case of Schlögl's model, $C^3 y = 0$ is just a shorthand of Eq. (6). If $d_a > 1$ the above equations form an underdetermined system of equations. In the following, we constrain the set of solutions with the use of so-called moment inequalities.

2. Semidefinite inequalities for the moments

The rational moments of any probability distribution on the nonnegative integers satisfy certain semidefinite inequalities^{43,44}. In particular, we define the *moment matrices*

$$\begin{aligned} (M_d^0(y))_{\alpha\beta} &:= y_{\alpha+\beta}, & \forall \alpha, \beta \in \mathbb{N}_{\lfloor d/2 \rfloor}^S, \\ (M_d^i(y))_{\alpha\beta} &:= y_{\alpha+\beta+e_i}, & \forall \alpha, \beta \in \mathbb{N}_{\lfloor (d-1)/2 \rfloor}^S, \end{aligned}$$

for $i = 1, \dots, S$, where $\mathbb{N}_d^S := \{n \in \mathbb{N}^S : |n| \leq d\}$, and e_i denotes the i^{th} unit vector. For instance, $M_3^0(y)$ and $M_3^1(y)$ are the matrices we employed in (7) for Schlögl's model. If y is the vector of rational moments of some distribution p on \mathbb{N}^S , that is

$$y_\alpha := \left\langle \frac{n^\alpha}{q(n)} \right\rangle_p, \quad \forall \alpha \in \mathbb{N}_d^S, \quad (12)$$

then these matrices are positive semidefinite (or in short $M_d^0(y) \succeq 0$, $M_d^i(y) \succeq 0$). This follows from the fact that for any polynomial $g(n)$ of degree $\lfloor d/2 \rfloor$ with polynomial coefficients g_α , it holds that

$$\begin{aligned} g^T M_d^0(y) g &= \sum_{|\alpha| \leq \lfloor d/2 \rfloor} \sum_{|\beta| \leq \lfloor d/2 \rfloor} g_\alpha g_\beta y_{\alpha+\beta} \\ &= \left\langle \frac{1}{q(n)} \left(\sum_{|\alpha| \leq \lfloor d/2 \rfloor} g_\alpha n^\alpha \right) \left(\sum_{|\beta| \leq \lfloor d/2 \rfloor} g_\beta n^\beta \right) \right\rangle_p \\ &= \left\langle \frac{g^2(n)}{q(n)} \right\rangle_p \geq 0, \end{aligned}$$

since $q(n) > 0$. Similarly, we have

$$g^T M_d^i(y) g = \left\langle \frac{n_i g^2(n)}{q(n)} \right\rangle_p \geq 0.$$

Additionally, since any probability distribution must have mass of one

$$q^T y = \sum_{|\alpha| \leq d_q} q_\alpha \left\langle \frac{n^\alpha}{q(n)} \right\rangle_p = 1.$$

3. Bounding moments using semidefinite programs

In summary, the set of vectors satisfying both the rational moment equations and the positive semidefinite inequalities is a spectrahedron, a convex set defined by matrix inequalities, which reads

$$\mathcal{E}_q^d := \left\{ y \in \mathbb{R}^{\#_d} : \begin{array}{l} q^T y = 1 \\ C^d y = 0 \\ M_d^i(y) \succeq 0, \quad \forall i = 0, 1, \dots, S \end{array} \right\}, \quad (13)$$

where $\#_d$ denotes the cardinality of \mathbb{N}_d^S . Denoting the set of stationary distributions with rational moments up to order d by $\mathcal{P}^{d,q}$, that is the set of $p \in \mathcal{P}$ such that all the averages in Eq. (10) are finite for all $|\alpha| \leq d$, we obtain the following straightforward theorem:

Theorem 1. *For any p in $\mathcal{P}^{d+1,q}$, the vector containing the rational moments of order d or less of p (defined in Eq. (12)) belongs to \mathcal{E}_q^d .*

A proof of the theorem can be found in Appendix B. This result has the following useful consequences that for any polynomial $f(n)$ of degree d and $p \in \mathcal{P}^{d,q}$

$$\hat{l}_f^d \leq \left\langle \frac{f(n)}{q(n)} \right\rangle_p \leq \hat{u}_f^d,$$

where

$$\hat{l}_f^d = \inf\{f^T y : y \in \mathcal{E}_q^d\}, \quad \hat{u}_f^d = \sup\{f^T y : y \in \mathcal{E}_q^d\}. \quad (14)$$

In particular, choosing $f(n) := q(n)n^\alpha$ with $|\alpha| \leq d - d_q$, it follows that the power moment $\langle n^\alpha \rangle_p$ is contained in the interval $[\hat{l}_f^d, \hat{u}_f^d]$ and thus we obtain the bounds of Eq. (11). Computing either \hat{l}_f^d or \hat{u}_f^d consists of solving the corresponding SDP given in Eq. (14).

Our result does not allow us to conclude on the convergence of these bounds with increasing number of moment equations and inequalities. However, it follows by construction that the sequence of lower bounds is non-decreasing while the sequence of upper bounds is non-increasing with increasing number of moments equalities and inequalities, i.e. it holds that

$$\hat{l}_f^d \leq \hat{l}_f^{d+1} \leq \dots, \quad \hat{u}_f^d \geq \hat{u}_f^{d+1} \geq \dots$$

In practice, however, we find that the bounds often converge to the true moments (or, in the case of multiple stationary distributions, the maximum/minimum moments

over the set of stationary distributions). For example, for Schlögl's model, in Fig. 1b we can see the projection of \mathcal{E}_1^d onto the y_1 - y_2 plane collapsing onto the point $(\langle n \rangle_p, \langle n^2 \rangle_p)$ as d is increased.

To investigate this convergence further, in Fig. 2a we show the upper and lower bounds on the first three moments as a function of d for Schlögl's model with a different parameter set than that in Fig. 1b. The insets show that the difference between upper and lower bounds decreases with d indicating that either bound converges to the true moment. Similarly, using these bounds we obtain converging bounds on several commonly used summary statistics as shown in Fig. 2b.

C. Implementation details and numerical considerations

To set-up the SDPs we used the modelling package YALMIP⁴⁵, and to solve them we used the solver SDPA-GMP⁴⁶ in conjunction with the interface mpYALMIP⁴⁷.

Before proceeding onto bounding the distributions, we should mention that SDPs associated with moment problems, such as those we have discussed in this section, are often ill conditioned. The reasons why are as of yet unclear. As we discuss in Section 5 of³², we believe they revolve around the fact that the moments of distributions change order of magnitude very quickly which leads to numerical instability in the solvers. This, and the success encountered in⁴⁸ when using SDP-GMP to tackle similar SDPs, motivated us to employ the multi-precision solver SDP-GMP instead of a standard double-precision SDP solver. Alternatively, employing rational moments, even when the network has polynomial propensities, can ameliorate this problem.

IV. BOUNDING THE STATIONARY DISTRIBUTIONS

Here we explain how to bound the probability that the stationary distributions give to any given subset of \mathbb{N}^S , and, in the case of a unique stationary distribution, how to iterate our approach to yield informative bounds on the complete distribution and its marginals. We begin by illustrating our approach with analytically tractable one-step processes. We then present our main result which enables us to use linear programming to bound the probability that the distributions give to any given subset of \mathbb{N}^S . The case of a unique stationary distribution is considered next. Lastly, we consider the non-unique case and we show how our method provides a computational test for uniqueness of the stationary distribution and how to adapt it to yield informative bounds in the non-unique case.

A. Semi-analytical bounds for one-step processes

For simplicity, we first focus on the case of a one-step process that involves a single species whose molecule count changes in the reactions by ± 1 . In this case, the stationary CME (2) can be solved recursively¹⁴ to obtain

$$p_n = \prod_{i=1}^n \underbrace{\frac{a_+(i-1)}{a_-(i)}}_{=:h_n} p_0, \quad n = 1, 2, \dots, \quad (15)$$

where $a_{\pm}(n)$ are the propensities of the forward and backward jumps and we have defined the function h_n . The probability p_0 is obtained by invoking the normalising condition

$$1 = p_0 \sum_{n=0}^{\infty} h_n. \quad (16)$$

Obtaining p_0 analytically is, however, possible only in simple cases.

1. Upper bounds

In cases for which computing p_0 analytically is not possible, it is numerically straightforward to truncate the sum in Eq. (16) considering only states for which $n < r$. While this procedure may be commonly used in practice, it does only provide an upper bound on the probability distribution. Specifically, since $1 \geq p_0 \sum_{n=0}^{r-1} h_n$, for all $r = 1, \dots$, we have that

$$p_n \leq \frac{h_n}{\sum_{n=0}^{r-1} h_n} =: u_n^r, \quad (17)$$

where the right hand side, u_n^r , is an upper on p_n for every $n < r$. Clearly, u_n^r tends to p_n as $r \rightarrow \infty$.

2. Lower bounds

Less obvious, however, is how to obtain lower bounds on p_n . To this end, we consider the mass μ_r of the tail of the distribution that can be bounded using Markov's inequality as follows

$$\mu_r := \sum_{n=r}^{\infty} p_n \leq \frac{\langle n^\alpha \rangle_p}{r^\alpha}.$$

However, explicit expressions for the involved moment are available only for linear propensities. To this end, we employ an upper bound $\langle n^\alpha \rangle_p \leq \hat{u}_\alpha^d$ that is readily computable using the method discussed in Sec. III. This provides us with a computable bound on the mass of the tail

$$\mu_r \leq \frac{\hat{u}_\alpha^d}{r^\alpha} =: \varepsilon_r.$$

Since

$$p_0 \sum_{n=0}^{r-1} h_n = \sum_{n=0}^{r-1} p_n = 1 - \mu^r \geq 1 - \varepsilon_r,$$

and using Eqs. (16,17) we obtain the lower bound

$$p_n = h_n p_0 \geq u_n^r (1 - \varepsilon_r) =: l_n^r.$$

3. Error bounds

Generally, we would like to determine how close the vector of upper bounds $u^r := (u_n^r)_{n=0,1,\dots,r-1}$ is to the vector composed of the first few values of the stationary distribution $p^r := (p_n)_{n=0,1,\dots,r-1}$ for a given r . To this end, we consider their ℓ^1 -distance (or, equivalently, twice the total variation distance)

$$\|u^r - p^r\|_1 = \sum_{n=0}^{r-1} (u_n^r - p_n) = \mu_r \leq \varepsilon_r, \quad (18)$$

where we have used the fact that $\sum_{n=0}^{r-1} u_n^r = 1$ by definition (17).

Similarly, the ℓ^1 -distance between the vector of lower bounds $l^r := (l_n^r)_{n=0,1,\dots,r-1}$ and p^r is bounded by

$$\|p^r - l^r\|_1 = \sum_{n=0}^{r-1} (p_n - l_n^r) = \varepsilon_r - \mu_r \leq \varepsilon_r. \quad (19)$$

Alternatively, by combining the upper bound u_n^r on p_n the lower bound l_n^r we can quantify the remaining uncertainty in our knowledge of the individual probabilities of the stationary distribution:

$$\max\{|p_n - u_n^r|, |p_n - l_n^r|\} \leq u_n^r - l_n^r = \varepsilon_r u_n^r \leq \varepsilon_r. \quad (20)$$

Since ε_r tends to 0 as $r \rightarrow \infty$, Eqs. (18-20) imply that the vectors of bounds converge both statewise and in ℓ^1 -norm to p as $r \rightarrow \infty$.

4. Schlögl's model

As an application of these distribution bounds we consider Schlögl's model introduced in Sec. III A. The first and third propensities in Eq. (5) can be lumped together into $a_+(n) := a_1(n) + a_3(n)$, while the second and fourth reaction can be lumped together into $a_-(n) := a_2(n) + a_4(n)$. The model enjoys the particular advantage that it can be solved exactly and it thus provides an ideal test case for the derived bounds. Using the normalising condition and Eq. (16), p_0 can be expressed in terms of a generalised hypergeometric function

$$\frac{1}{p_0} = {}_2F_2 \left(-\frac{c_1+1}{2}, \frac{c_1-1}{2}; -\frac{c_2+1}{2}, \frac{c_2-1}{2}; \frac{k_1}{k_2} \right), \quad (21)$$

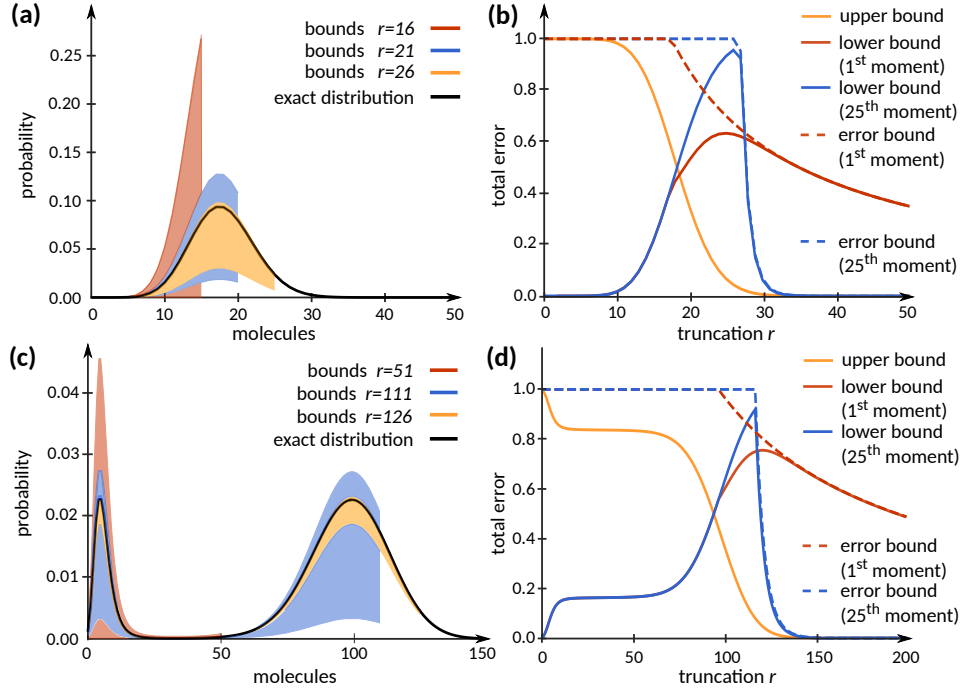


FIG. 3. **Bounding the probability distribution of a one-step process.** Schlögl's model Eqs. (4)-(5) is an example of a one-step process which has both unimodal (a,b) and bimodal (c,d) stationary distributions depending on kinetic parameters. (a) Parameters resulting in a unimodal distribution. Shaded areas between upper and lower bounds on the probability distributions are shown using the first $r = 16$ (red), $r = 21$ (green) and $r = 26$ (blue) stationary equations obtained using a bound on the first moment. The black line shows the exact distribution, Eqs. (15) and (21). The lower bounds were all computed using the upper bound on the 1st moment. (b) The total error computed using the ℓ^1 -distance between the exact solution and the upper bound (yellow) decreasing with increasing truncation r . The total error of the lower bounds is computed using a bound on the 1st moment (red) and 25th moment (blue line). Error bounds given by Eq. (19) and (20) are in good agreement with the actual errors for moderate to large truncations. (c) Bounds for different parameters yielding a bimodal distribution for $r = 51$ (red, scaled by a factor of $1/3$), $r = 111$ (green), $r = 126$ (blue) are compared to the exact distribution (black). The lower bounds were computed using the upper bound on the 1st moment for $r = 51, 126$ and that on the 25th moment for $r = 126$. (d) Total errors and error bounds shown as in (b). In (a) and (b) parameter values are $k_1 = 6$, $k_2 = 1/3$, $k_3 = 50$, $k_4 = 3$ yielding upper bounds of $\hat{u}_1^{25} = 17.5$ and $\hat{u}_{25}^{25} = 5.79 \times 10^{35}$ on the 1st and 25th moment, respectively. In (c) and (d) we used $k_1 = 1/9$, $k_2 = 1/1215$, $k_3 = 27/2$, $k_4 = 59/20$ which gives upper bounds of $\hat{u}_1^{25} = 98.0$ and $\hat{u}_{25}^{25} = 6.37 \times 10^{51}$.

with $c_1 = \sqrt{1 - 4k_3/k_1}$ and $c_2 = \sqrt{1 - 4k_4/k_2}$. To compute the bounds, we require a bound on the mass of the tail. To this end, we compute upper bounds on the 1st and 25th moments using the first 23 moment equations and the corresponding matrix inequalities involving the first 25 moments. We considered two parameter sets, one for which the stationary distribution is unimodal, and another for which it is bimodal. In Figs. 3a and 3c, we compare the bounds we obtained with the exact distribution computed using Eq. (15) and (21).

In practice, we need to choose both the truncation parameter r and which moment bound to employ when computing the vectors l^r and u^r . As Lemmas 3 and 4 in the following section show, we can always use the computed vectors and the moment bound to a posteriori bound the ℓ^1 -distance between the vector of bounds and the stationary distribution, which we refer to as the “total error”. If the error bound is unsatisfactory, then we can increase r and re-compute these bounds. Choosing a good r and moment bound combination a priori is not so

straightforward. A basic guideline follows from Jensen's inequality

$$\varepsilon_r = \frac{\hat{u}_\alpha^d}{r^\alpha} \geq \frac{\langle n^\alpha \rangle_p}{r^\alpha} \geq \left(\frac{\langle n \rangle_p}{r} \right)^\alpha.$$

In words, to achieve an error of less than $1/2$, r must always be greater than the mean number of molecules. To investigate this matter further, we studied in Figs. 3b and 3d how the error and the bound on the error depend on r and on which moment we employed for Schlögl's model. The error (solid lines) was computed using the analytical expression available for the stationary distribution of this simple network, Eqs. (15) and (21). For more general networks where no such expression is known, the error is not computable. However, the error bounds (dashed lines) are still readily available as their computation only involves the vectors u^r , l^r , and the bound on the moment. From the definition of the error bound ε_r , it is clear that the error bound converges to zero far quicker when using the 25th moment (blue dashed line) rather

than the 1st moment (red dashed line), in good agreement with the actual error (solid blue and red). Hence for large truncations r it is preferable to use a higher moment to compute the distribution bounds. However, for moderate r it is interesting to note that lower moments can also outperform higher ones.

B. Linear programming approach to multi-species networks

The approach of the previous section relied on expression (2), only available for single-species one-step process, that relates the whole stationary distribution p to one single entry p_0 . To generalise our approach to multi-species networks, we deal directly the stationary equations. Specifically, we consider the equations for $n := (n_1, n_2, \dots, n_S)$ such that

$$|n|_w := w_1 n_1 + w_2 n_2 + \dots + w_S n_S < r, \quad (22)$$

where the w_i 's are positive weights. Introducing such weights is important in applications because it allows choosing truncations based on the abundance of different species. There are $\#_r^w$ -many such equations involving the $\#_r^w$ -many states contained in

$$\mathcal{N}_r := \{n \in \mathbb{N}^S : |n|_w < r\} \\ \bigcup_{j=1}^R \{n + v^j \in \mathbb{N}^S : |n|_w < r\}. \quad (23)$$

Let \mathbb{R}^∞ denote the space of all real-valued vectors indexed by \mathbb{N}^S and P_r (resp. \tilde{P}_r) be the projection operator that takes a vector $z \in \mathbb{R}^\infty$ and returns $(z_n)_{\{n: |n|_w < r\}} \in \mathbb{R}^{\#_r^w}$ (resp. $(z_n)_{n \in \mathcal{N}_r} \in \mathbb{R}^{\#_r^w}$). We can write these first few equations as $Q_r \tilde{P}_r p = 0$, where $Q_r := P_r Q \tilde{P}_r^T$ is a $\#_r^w \times \#_r^w$ -dimensional matrix. These equations are underdetermined since, except for finite state spaces, $\#_r^w < \#_r^w$. We overcome this issue in a similar manner to how we overcame the issue for the moments: we supplement these equations with inequalities and then compute the bounds by solving LPs (a subclass of SDPs). To this end, we require an upper bound on a moment, which via Markov's inequality also equips us with an upper bound on the mass of the tail.

1. Bounding event probabilities

Let \mathcal{P}_c^d denote the subset of stationary distributions p such that $\langle |n|_w^d \rangle_p \leq c$, where c is a constant. A bound on the mass of the tail μ_r , i.e. the mass in the complement of \mathcal{N}_r , is obtained as follows

$$\mu_r := \sum_{n \in (\mathcal{N}_r)^c} p_n \leq \sum_{|n|_w \geq r} p_n \leq \frac{\langle |n|_w^d \rangle_p}{r^d} \leq \frac{c}{r^d}, \quad (24)$$

for any $p \in \mathcal{P}_c^d$ and the summation in the second sum is over all states n for which $|n|_w \geq r$. The second inequality is an instance of Markov's inequality. Consider now the convex polytope defined by

$$\mathcal{B}^r := \left\{ z \in \mathbb{R}^{\#_r^w} : \begin{array}{l} z \geq 0, \\ 1 - c/r^d \leq 1^T z \leq 1, \\ Q_r z = 0, \\ \sum_{n \in \mathcal{N}_r} |n|_w^d z_n \leq c \end{array} \right\}.$$

For any $A \subseteq \mathbb{N}^S$ (resp. $A \subseteq \mathcal{N}_r$) let $1_A \in \mathbb{R}^\infty$ (resp. $1_A \in \mathbb{R}^{\#_r^w}$) to denote the indicator vector

$$(1_A)_n := \begin{cases} 1 & \text{if } n \in A \\ 0 & \text{otherwise} \end{cases},$$

for all $n \in \mathbb{N}^S$ (resp. $n \in \mathcal{N}_r$). For any p , the probability of an event A is then given by $1_A^T p$, and can be bounded from above and below as follows:

Theorem 2. *Pick any $A \subseteq \mathbb{N}^S$, let $A^r := A \cap \mathcal{N}_r$ and*

$$l_A^r := \inf\{1_{A^r}^T z : z \in \mathcal{B}^r\}, \quad u_A^r := \sup\{1_{A^r}^T z : z \in \mathcal{B}^r\}.$$

1. *For any $p \in \mathcal{P}_c^d$, we have that*

$$l_A^r \leq 1_{A^r}^T p, \quad 1_{A^r}^T p \leq u_A^r$$

for all $r \geq 0$.

2. *Suppose that \mathcal{P}_c^d is non-empty and let*

$$l_A := \inf\{1_A^T p : p \in \mathcal{P}_c^d\}, \quad u_A := \sup\{1_A^T p : p \in \mathcal{P}_c^d\}.$$

Both sequences $\{l_A^r\}_{r \geq 0}$ and $\{u_A^r\}_{r \geq 0}$ converge and

$$\lim_{r \rightarrow \infty} l_A^r = l_A, \quad \lim_{r \rightarrow \infty} u_A^r = u_A.$$

A proof of the above theorem is given in Appendix C. Note that by definition $1_{A^r}^T p \leq 1_{A^r}^T p$ and thus l_A^r is indeed a lower bound on $1_{A^r}^T p$. The number u_A^r will be an upper bound on $1_{A^r}^T p$ if $A \subseteq \mathcal{N}_r$. Computing either l_A^r or u_A^r consists of solving a LP, that can efficiently be done using any one of the solvers are readily available online. The practical implications of the theorem are discussed in the following sections.

2. Unique case: Bounding probability distributions

If there exists a unique stationary distribution p , we are often interested in computing the complete distribution instead of the probability that p assigns to a single subset A of the state space \mathbb{N}^S . Suppose that $\mathcal{P}_c^d = \{p\}$ and $A = \{n\}$, then Theorem 2 implies that

$$l_n^r := l_{\{n\}}^r \leq p_n \leq u_{\{n\}}^r := u_n^r,$$

statewise for each $n \in \mathcal{N}_r$. Additionally, we have that both l_n^r and u_n^r converge to p_n as $r \rightarrow \infty$.

Denoting by $l^r := (l_n^r)_{n \in \mathcal{N}_r}$ the vector of lower bounds for all states $n \in \mathcal{N}_r$ and similarly for u^r , it follows that

$$\max \left\{ \left\| \tilde{P}_r p - l^r \right\|_1, \left\| \tilde{P}_r p - u^r \right\|_1 \right\} \leq \|u^r - l^r\|_1.$$

In words, by computing the lower bounds in l^r and the upper bounds in u^r we can bound the ℓ^1 -distance between p and either vector of bounds. We refer to this distance as the “total error” of the vector of bounds and we refer to any bound on the total error, such as that appearing in the above, as an “error bound”. Additionally, we have the following two other error bounds

Lemma 3. *For any $p \in \mathcal{P}_c^d$, we have that*

$$\left\| \tilde{P}_r p - l^r \right\|_1 = 1 - 1^T l^r - \mu^r \leq 1 - 1^T l^r, \quad (25a)$$

$$\left\| u^r - \tilde{P}_r p \right\|_1 = 1^T u^r - 1 + \mu^r \leq 1^T u^r - 1 + \frac{c}{r^d}. \quad (25b)$$

That is, we can also bound the total error of one vector of bounds without computing the other vector of bounds. A proof of the above lemma can be found in Appendix C.

3. Unique case: Bounding marginal distributions

For high-dimensional networks we are often interested in marginal distributions rather than the full stationary distribution. For simplicity, we assume that we are interested only in the first \tilde{S} species, so that we are marginalising over the last $\hat{S} := S - \tilde{S}$ species. We decompose the vectors of molecule numbers into (n, m) where $n \in \mathbb{N}^{\tilde{S}}$ and $m \in \mathbb{N}^{\hat{S}}$ such that the marginal distribution \tilde{p} on $\mathbb{N}^{\tilde{S}}$ is as follows

$$\tilde{p}_n := \sum_{m \in \mathbb{N}^{\hat{S}}} p_{(n, m)} \quad \forall n \in \mathbb{N}^{\tilde{S}}.$$

To proceed we introduce the n -slice of \mathcal{N}_r in Eq. (23),

$$\mathcal{N}_{r, n} := \{m \in \mathbb{N}^{\hat{S}} : (n, m) \in \mathcal{N}_r\},$$

and let $A := \mathcal{N}_{r, n}$ in Theorem 2. Thus, for each $n \in \mathbb{N}^{\tilde{S}}$ for which there exists at least one $m \in \mathbb{N}^{\hat{S}}$ such that $(n, m) \in \mathcal{N}_r$, we find that

$$\tilde{l}_n^r := \inf \left\{ \sum_{m \in \mathcal{N}_{r, n}} z_{(n, m)} : z \in \mathcal{B}^r \right\},$$

is a lower bound on \tilde{p}_n . In contrast,

$$\tilde{u}_n^r := \sup \left\{ \sum_{m \in \mathcal{N}_{r, n}} z_{(n, m)} : z \in \mathcal{B}^r \right\}, \quad (26)$$

is not necessarily an upper bound for \tilde{p}_n because marginalising over m involves infinitely many states not included in the truncated space $\mathcal{N}_{r, n}$. However, the probability mass of these states is bounded by the mass of the tail μ_r and hence by c/r^d in Eq. (24). It follows that $\tilde{u}_n^r + c/r^d$ is an upper bound on \tilde{p}_n . We thus have that

$$\tilde{l}_n^r \leq \tilde{p}_n \leq \tilde{u}_n^r + \frac{c}{r^d} =: \tilde{u}_n^{*r},$$

for each $n \in \mathbb{N}^{\tilde{S}}$ for which there exists an $m \in \mathbb{N}^{\hat{S}}$ such that $(n, m) \in \mathcal{N}_r$. Moreover, Theorem 2 implies that both \tilde{l}_n^r and \tilde{u}_n^r converge to \tilde{p}_n as $r \rightarrow \infty$ (and, consequently, so does \tilde{u}_n^{*r}). As an analogue of Lemma 3, the following result gives error bounds on the marginal distribution estimates:

Lemma 4. *For any $p \in \mathcal{P}_c^d$, we have that*

$$\left\| \tilde{P}_r \tilde{p} - \tilde{l}^r \right\|_1 \leq 1 - 1^T \tilde{l}^r, \quad (27a)$$

$$\left\| \tilde{u}^r - \tilde{P}_r \tilde{p} \right\|_1 \leq 1^T \tilde{u}^r - 1 + \frac{2c}{r^d}, \quad (27b)$$

$$\left\| \tilde{u}^{*r} - \tilde{P}_r \tilde{p} \right\|_1 \leq 1^T \tilde{u}^{*r} - 1 + \frac{c}{r^d}. \quad (27c)$$

A proof of the above lemma can be found in Appendix C.

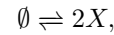
4. Non-unique case

In the case where the network admits multiple stationary distributions, our method provides lower and upper bounds on the set of all stationary distributions. That is, setting $A := \{n\}$ in Theorem 2, we obtain the bounds

$$l_n^r \leq \inf \{p_n : p \in \mathcal{P}_c^d\} \leq p_n \leq \sup \{p_n : p \in \mathcal{P}_c^d\} \leq u_n^r,$$

that hold for any $p \in \mathcal{P}_c^d$.

Non-uniqueness occurs due to different stationary distributions being reached from different initial conditions (these stationary distributions are known as the ergodic distributions). For example, the network



modelled using mass action propensities admits two ergodic distributions depending on whether we start the network with an even or odd number of molecules. Each of these distributions has support on distinct irreducible subsets of the state space, namely the even or odd natural numbers. The set of stationary distributions is the set of convex combinations of the ergodic distribution¹². If n is even (resp. odd), the ergodic distribution corresponding to the odd (resp. even) numbers assigns zero probability to state n . Thus $\inf \{p_n : p \in \mathcal{P}_c^d\} = 0$ and it follows that are $l_n^r = 0$ for all r such that $n \in \mathcal{N}_r$.

Conversely, by the above reasoning, if, for any single n , our method returns a single non-zero lower bound l_n^r ,

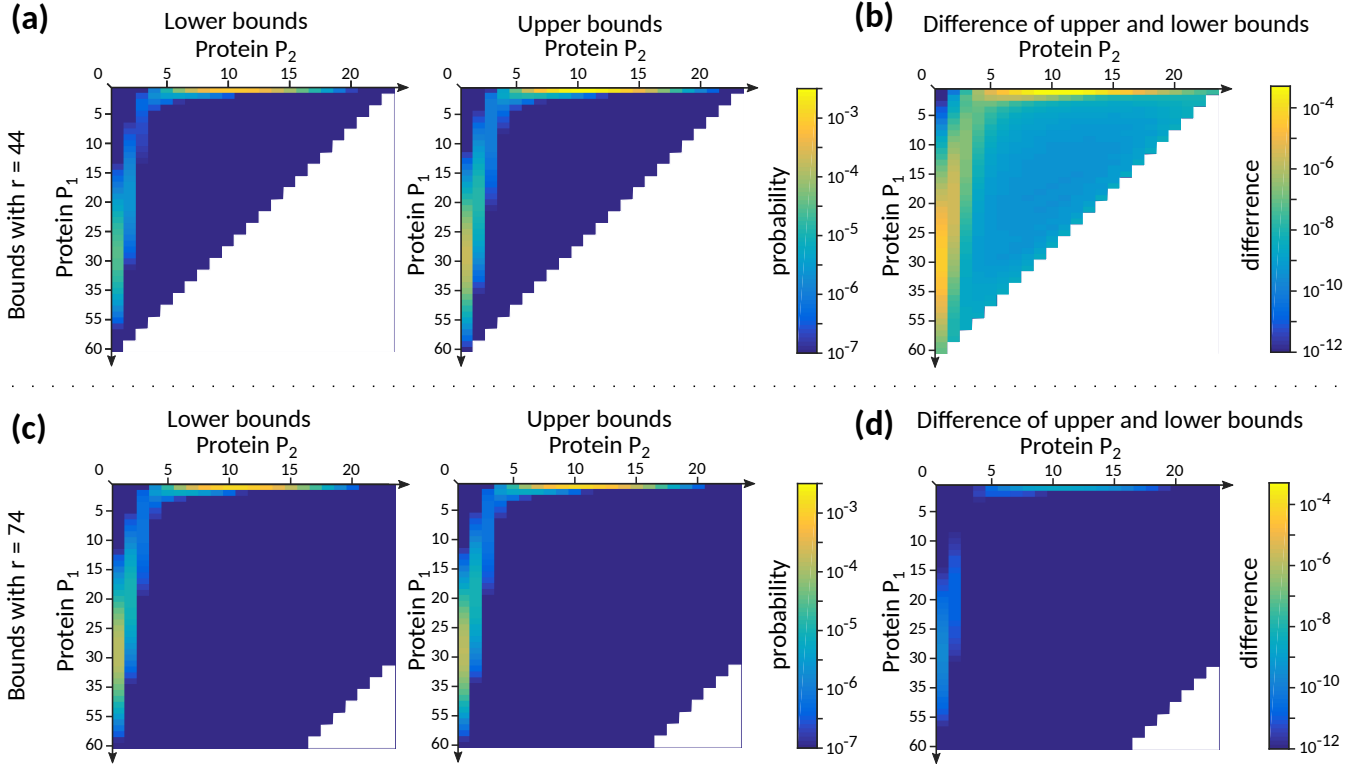


FIG. 4. **Bounds on the stationary probability of the toggle switch.** (a) Upper lower and bounds are shown to give qualitatively similar bimodal distribution shapes. The state truncation parameter was chosen $r = 44$, so that 529 equations involving 575 states were included. White areas denote states n for which $|n|_w > 44$ and that our estimate contains no information about. Computation time was 2.2 minutes on a desktop computer. (b) Computing the difference between these bounds reveals that the largest uncertainties occur near the distribution modes. Note the difference in the values of the colourmap with those in the colourmap in (a). (c) The state truncation parameter was refined to $r = 74$, so that 1444 equations involving 1520 states were included. Upper and lower bounds are visually indistinguishable. Computation time was 22.6 minutes on a desktop computer. (d) Computing the difference of these bounds reveals that these bounds are close to the true distribution with the maximum absolute difference being as small as 10^{-9} . Parameters are given in the main text.

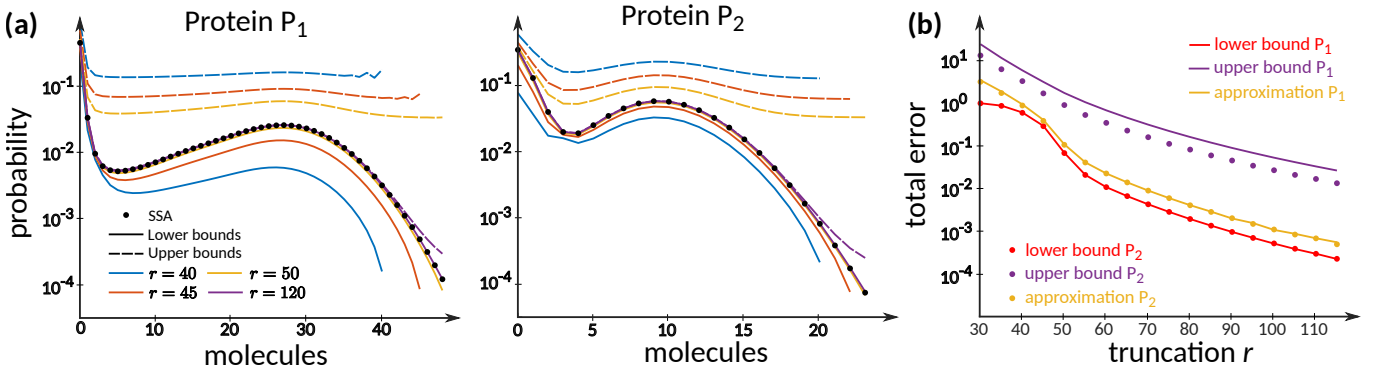


FIG. 5. **Bounds on the marginal probability distributions of the toggle switch.** (a) Lower bounds \tilde{l}^r (solid lines) and upper bounds \tilde{u}^{*r} (dashed lines) on the marginal distributions for increasing state space truncations $r = 40$ (blue), 45 (red), 50 (yellow) and 120 (purple). Both sets of bounds converge to the stationary distribution of the CME. For comparison, we show simulations performed using the stochastic simulation algorithm (SSA, black dots). (b) Error bounds for the lower (red, Eq. 27a) and upper bounds (purple, Eq. 27c). We also show the error obtained by using the approximation u^r (yellow, Eq. (27b)). Lines show the error bounds obtained for the marginal distribution of protein P_1 , while dots show those the corresponding to the marginal distribution of protein P_2 . The total error of all estimates decreases with the number of states considered (r). Parameters as in Fig. 4.

then there must exist at most a single stationary distribution. While uniqueness is typically verified by checking irreducibility of the state-space^{40,49}, our method provides a direct computational criterion to establish uniqueness of the stationary distribution. This criterion is not only sufficient but necessary in the sense that, by the convergence of the bounds, if there is a unique stationary distribution p , then l_n^r will be positive for large enough r and n such that $p_n > 0$.

If non-uniqueness holds but the irreducible sets are known, like in the above example, our method can be modified to provide informative bounds on the ergodic distributions. In particular, as can be verified from the proofs in Appendix C, given any support set \mathcal{S} , replacing \mathcal{B}^r with

$$\mathcal{B}_S^r := \mathcal{B}^r \cap \{z \in \mathbb{R}^{\#_r^w} : z_n = 0, \quad \forall n \notin \mathcal{S} \cap \mathcal{N}_r\},$$

then Theorem 2, Lemma 3, and 4 hold with \mathcal{P}_c^d replaced by

$$\mathcal{P}_c^{d,\mathcal{S}} := \mathcal{P}_c^d \cap \{p \in \mathbb{R}^\infty : p_n = 0, \quad \forall n \notin \mathcal{S}\},$$

the set of stationary distributions with support contained in \mathcal{S} . So, for instance, if we pick \mathcal{S} to be the even (resp. odd) natural numbers in the above example, our bounds converge to the network's ergodic distribution on the even (resp. odd) natural numbers.

V. APPLICATION: A TOGGLE SWITCH

As an application of the methods discussed in this paper, we consider a network of two mutually repressing genes. Such toggle switches are common motifs in many cell fate decisions^{13,50,51}. In particular, we consider the asymmetric case

$$\emptyset \xrightleftharpoons[a_2(n)]{a_1(n)} P_1, \quad \emptyset \xrightleftharpoons[a_4(n)]{a_3(n)} P_2,$$

whose propensities are

$$\begin{aligned} a_1(n) &= \frac{k_1 \theta^3}{\theta^3 + n_{P_2}^3}, \quad a_2(n) = k_2 n_{P_1}, \\ a_3(n) &= \frac{k_3}{1 + n_{P_1}}, \quad a_4(n) = k_4 n_{P_2}, \end{aligned} \quad (28)$$

which model mutual repression via Hill functions and dilution via linear unspecific decay. The model has a unique stationary distribution and all its moments are finite for all sets of positive kinetic parameters, see Appendix A.

A. Bounds on the joint distribution

For constructing a tail bound, we require an upper bound on a certain moment. Since our example involves

rational propensities, we use rational moments as defined in Eq. (10) with

$$q(n) = (1 + n_{P_1}) (1 + (n_{P_2}/\theta)^3),$$

given by the common denominator of the propensities, Eq. (28). Due to asymmetric repression of the two proteins, the range of molecule numbers for protein P_1 is larger than for protein P_2 , which we account for using the weight vector $w := (1, 2)$ in the state space truncation, Eq. (22). We then compute an upper bound on the 6th moment of

$$\langle |n|_w^6 \rangle_p \leq c = 5.0901 \times 10^8,$$

using the SDP (11) with \mathcal{E}_q^{10} in Eq. (13) including the first $d = 10$ rational moments and the parameters $\theta = 1$, $k_1 = 30$, $k_2 = k_4 = 1$ and $k_3 = 10$. Markov's inequality implies that the mass of the tail μ_r is bounded by

$$\mu_r \leq \frac{c}{r^6},$$

depending on the state space truncation parameter r .

Using the method described in Sec. IV B 2, we computed upper and lower bounds on the joint distributions for small ($r = 44$) and large ($r = 74$) state space truncations as shown in Fig. 4. These bounds give an accurate account of the uncertainty in the probabilities using only the first few states of the CME. For small state spaces, we find that the maximum absolute discrepancies are found near the distribution modes (Fig. 4a,b). Increasing the number of states, the upper and lower bounds are nearly indistinguishable (Fig. 4c), the difference between these bounds is smaller than 10^{-9} (Fig. 4d).

B. Bounds on the marginal distributions

The method described in Sec. IV B 3 allows us to obtain upper and lower bounds on the marginal distributions as well. The corresponding bounds for the distributions of protein P_1 and P_2 are shown in Fig. 5a and demonstrate the convergence of our estimates. To verify our implementation we compare with histograms obtained from stochastic simulations. In Fig. 5b we show the corresponding total error estimates given by Eq. (27a) and (27c) that bound the distance between the respective lower and upper bounds and the underlying marginal distribution. We also show the error estimate on the upper bound of the truncated state space (yellow), which does not provide an upper bound on the marginal distributions but rather presents an approximation.

Finally, we apply the method to gain insights over whole parameter ranges. For example, increasing the dissociation constant θ of protein P_2 from zero to finite values allows to induce expression of protein P_1 . To quantify this dependence we use the lower bounds on the marginal distributions shown in Fig. 6. For intermediate levels of θ ,

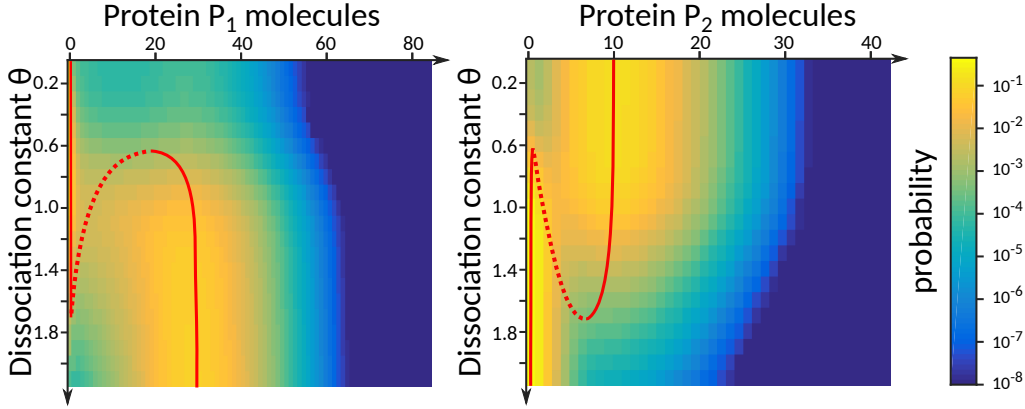


FIG. 6. **Flipping the toggle switch: deterministic vs. stochastic models.** Tuning the promoter dissociation constant θ from small to high values results in a switch-like increase of the production of protein P_1 molecules while simultaneously repressing the production of protein P_2 . In deterministic models such tuning leads to two stable steady states in the transition region (solid red lines) accessible from different initial conditions. In the stochastic model the marginal stationary probabilities (heatmap) displays two coexisting populations. We observe good correspondence between the modes of the distributions and the deterministically stable steady states. Each marginal is obtained from lower bounds computed using $r = 83$ resulting in 1806 equations involving 1891 states. The error bound according to Eq. (27a) is at most 3×10^{-3} . Parameters as in the main text except for θ .

we observe that induction of P_1 does not occur gradually but is represented by two coexisting populations corresponding to fully induced and repressed populations, i.e. one with high levels of P_1 but low levels of P_2 and another one with low levels of P_1 but high levels of P_2 . We find that the modes of the marginal distributions are in good correspondence with the stable solutions of the steady-state rate equations.

C. Implementation details and numerical considerations

To set-up the LPs we used the modelling package YALMIP⁴⁵, and to solve them we used the solver CPLEX⁵². As in Section IIIB, for the SDPs we used YALMIP⁴⁵, SDP-GMP⁴⁶, and mpYALMIP⁴⁷.

Numerically, in the case of the LPs, the constraint relating to the moment bound makes the problem data ill conditioned and causes the solvers struggle. One can remove this constraint, the LPs still yield bounds, and the error bounds in Lemmas 3 and 4 still hold, as can be verified from the proofs in Appendix C. However, convergence of the bounds can no longer be established and, even though in our practical experience the bounds often still converge, we have encountered one example (a 4-D toggle switch) for which they appeared not to.

VI. CONCLUDING REMARKS

We developed an approach to bound stationary moments and distributions of multivariate biochemical reaction networks. These statistical quantities satisfy the stationary moment equations and the stationary CME, respectively, both which consist of a generally intractable

infinite set of equations. Considering a finite subset of these equations, the proposed method enables us to bound the set of solutions using mathematical programming techniques and yields accurate estimates of moments and probabilities with computable error bounds.

To compute the moment bounds, we supplement the moment equations with semidefinite inequalities that are satisfied by the moments of any stationary distribution and allow us to formulate a SDP whose solution yields an upper or lower bound on any given moment. Repeated applications of the method yield sequences of upper (resp. lower) bounds that are monotonically decreasing (resp. increasing) as a function with the number of moment equations and inequalities used (Theorem 1). Using biological relevant examples we demonstrate that these moment estimates often converge to the true moment in practice, although we have no theoretical proof of convergence.

To compute bounds on probabilities, we formulate a LP by supplementing the truncated CME with a bound on a single moment and one on the mass of the tail derived from the moment bound. Repeated applications of the method yield sequences of bounds that converge to the probability of interest (Theorem 2). We place no restrictions on the network nor its propensities, aside from the availability of a moment bound. In particular, in the case of a unique stationary distribution, we can use the method to obtain converging lower and upper bounds on the complete distribution and its marginals. We provide quantified error estimates for both set of bounds (Lemma 3), making our procedure a systematic self-contained approach to computing stationary distributions of reaction networks and their marginals. As an aside, we explained in Sec. IV B 4 how the method also provides a computational test for uniqueness of the stationary distributions,

and described a simple way to adapt it to the non-unique case.

An alternative approach to obtaining bounds on stationary probability distributions was been introduced by *Spieler et al.*^{53,54}. The method constructs a finite set of Markov chains whose minimum and maximum state-wise probabilities bound the stationary distributions. It does not rely on a moment bound but instead requires only a tail bound. Their approach to obtaining such a tail bound consists of finding a Lyapunov function on a case-by-case basis. A second approach for the class of networks with affine propensities was presented in⁵⁵. It also involves finding a Lyapunov function, but this time the search is carried out computationally. The problem of finding such Lyapunov functions is essentially dual⁵⁶ to that of finding the moment bounds as we do in Section III. Compared with these two other methods, our approach enjoys the advantages that it guarantees convergence of the bounds and that it is accompanied by computable error bounds.

In our experience, the disadvantages of our approach those of numerical nature discussed in Sections III C and V C. Future improvements in LP and SDP solvers likely will mitigate these issues, but at the present time they pose limitations to the applicability of the approaches discussed in this paper.

An interesting technical point that we have omitted from the main text is the following. When solving each LP one obtains a feasible vector $z^r \in \mathcal{B}^r$ that, up to some numerical tolerance, achieves the optimum (that is $1_{A^r}^T z^r = l_A^r$ or $1_{A^r}^T z^r = u_A^r$, depending on which LP we solved). If there is a unique stationary distribution p , then the sequence of vectors $\{z^r\}_{r \geq 0}$ converges pointwise to p , see Corollary C.2 1. in Appendix C. In other words, by solving a single LP for high r (instead of one for each state n) and extracting the optimal point z^r , we can obtain a good approximation of the stationary distribution, namely z^r . In the non-unique case, there is at most one stationary distribution p per positive recurrent state n such that $p_n = u_{\{n\}}$, namely the ergodic distribution with support on the irreducible set that n belongs to¹². Corollary C.2 2. shows that z^r converges to p . Thus, by solving a single LP for high r we can obtain a good approximation of p without any a priori knowledge of the regarding the irreducible set associated with p . The issue with these computationally cheaper approaches to approximating the stationary distribution, is that, as of yet, we know of no way to quantify the approximation error, that is we have no error bounds of the sort of those in Lemmas 3.

A distinct feature of our method is that it enables the direct computation of lower and upper bounds on the marginal distributions. In particular, we need not to compute bounds for each state of joint distribution, but only bounds for each state of the marginal distribution of interest. These distributions are of particular interest for the analysis of high-dimensional networks, for which we provide practical error estimates (Lemma 4). We thus

expect that our approach will be particularly valuable in applications where accurate approximations with quantified errors are needed. These include, for example, estimating distributions of phenotypic switches, determining persister fractions in bacterial populations and inferring model parameters from noisy single cell data⁵⁷.

VII. ACKNOWLEDGEMENTS

We would like to thank Justine Dattani and Diego Oyarzún for valuable discussions on the biological applications of CMEs and Michela Ottobre and Jure Vogrin for insightful discussions regarding the Markov chains underlying CMEs, their long-term behaviour, and the various relevant stability concepts. JK gratefully acknowledges support by the BBSRC in the form of the Ph.D studentship BB/F017510/1, PT by The Royal Commission for the Exhibition of 1851, GBS by the EPSRC Fellowship EP/M002187/1, and MB by the EPSRC grant number EP/N014529/1.

Appendix A: Existence, uniqueness, and finite of moments of stationary distributions

Theorems 2.1 and 4.2 in Ref.³⁹ show that least one stationary distribution exists and that all stationary distributions have finite d^{th} -order moment if there exists a function $u : \mathbb{N}^S \rightarrow [0, \infty)$ and constants $c_1, c_2 > 0$ such that

$$\begin{aligned} Q^T u(n) &\leq c_2 - c_1 |n|^d \quad \forall n \in \mathbb{N}^S, \\ \{n : \in \mathbb{N}^S : u(n) \leq c\} &\text{ is finite } \forall c \in \mathbb{R}. \end{aligned}$$

For Schögl's model we have that

$$Q^T n^{d-2} = g_{d-1}(n) - k_2(d-2)n^d,$$

where g_{d-1} is a polynomial of degree $d-1$. Thus, we have that

$$Q^T n^{d-2} \leq \sup_{n \in \mathbb{N}^S} \left\{ g_{d-1}(n) - \frac{k_2(d-2)}{2} n^d \right\} - \frac{k_2(d-2)}{2} n^d.$$

Picking $d \geq 3$ then tells us that at least one stationary distribution exists, and that $\langle n^1 \rangle_p, \dots, \langle n^d \rangle_p$ are finite for any stationary distribution p . Since we can choose ever larger d we have finiteness of all moments.

Similarly, to prove existence of stationary distributions and finiteness of all their moments for the toggle switch consider

$$\begin{aligned} Q^T |n|^d &= (a_1(n) + a_3(n))((n_{P_1} + n_{P_2} + 1)^d - (n_{P_1} + n_{P_2})^d) \\ &\quad + (a_2(n) + a_4(n))((n_{P_1} + n_{P_2} - 1)^d - (n_{P_1} + n_{P_2})^d) \\ &\leq (k_1 + k_3)((n_{P_1} + n_{P_2} + 1)^d - (n_{P_1} + n_{P_2})^d) \\ &\quad + (a_2(n) + a_4(n))((n_{P_1} + n_{P_2} - 1)^d - (n_{P_1} + n_{P_2})^d) \\ &\leq g_{d-1}(n_{P_1} + n_{P_2}) - d(k_2 n_{P_1} + k_4 n_{P_2})(n_{P_1} + n_{P_2})^{d-1} \\ &\leq g_{d-1}(n_{P_1} + n_{P_2}) - d \min\{k_2, k_4\} |n|^d \end{aligned}$$

The remainder of the argument is identical to that of Schögl's model.

Uniqueness follows from the fact that they are both one-step processes and consequently they are irreducible. Alternatively, for the particular parameter values we examined, uniqueness followed from the fact that we obtained non-zero lower bounds, see Figs. 3-4 and Section IV B 4.

Appendix B: Proof of Theorem 1

All that we must prove here is that Eq. (3) holds for $f(n) = n^\alpha$ for any $\alpha \in \mathbb{N}_{d-d_a+1}^S$ under the conditions in the theorem's premise. The theorem then follows directly from the discussion in Section III B and Eq. (3). We will show that Eq. (3) holds for any f such that

$$\sum_{j=1}^R \left(\langle a_j(n) | f(n) \rangle_p + \langle a_j(n) | f(n + v_j) \rangle_p \right) < \infty, \quad (\text{B1})$$

Then, if p has rational moments of order $d + 1$ (as in the theorem's premise), we have that Condition (B1) is satisfied for any $f(n) = n^\alpha$ with $\alpha \in \mathbb{N}_{d-d_a+1}^S$. Thus, Eq. (3) holds for any such f and we have that the rational moments satisfy $C^d y = 0$.

Noting that Condition (B1) implies that the integrals

$$\sum_{n \in \mathbb{N}^S} a_j(n) f(n + v_j), \quad \sum_{n \in \mathbb{N}^S} a_j(n) f(n), \quad \sum_{n \in \mathbb{N}^S} Q^T f(n)$$

are all well-defined and finite, the proof of Eq. (3) is straightforward:

$$\begin{aligned} \langle Q^T f \rangle_p &= \sum_{n \in \mathbb{N}^S} p_n \left(\sum_{j=1}^R a_j(n) (f(n + v_j) - f(n)) \right) \\ &= \sum_{j=1}^R \left(\left(\sum_{n \in \mathbb{N}^S} p_n a_j(n) f(n + v_j) - \sum_{n \in \mathbb{N}^S} p_n f(n) \right) \right) \\ &= \sum_{j=1}^R \left(\left(\sum_{n \in \mathbb{N}^S} p_{n-v_j} a_j(n - v_j) f(n) - \sum_{n \in \mathbb{N}^S} p_n f(n) \right) \right) \\ &= \sum_{n \in \mathbb{N}^S} \left(\sum_{j=1}^R (p_{n-v_j} a_j(n - v_j) - p_n) \right) f(n) \\ &= \sum_{n \in \mathbb{N}^S} (Qp)(n) f(n) = 0, \end{aligned}$$

where we have tacitly used the convention $p_n a_n := 0$ for any $n \notin \mathbb{N}^S$.

Condition (B1) is only sufficient for Eq. (3) to hold. A necessary condition is that $\langle |Q^T f(n)| \rangle_p < \infty$, otherwise Eq. (3) is ill-defined. However, for general chains, the weaker condition $\langle |Q^T f(n)| \rangle_p < \infty$ is not sufficient,

as Example 2 in Ref.⁵⁶ shows. We are unsure whether this is still the case for networks with rational propensity functions and polynomial f s. Fortunately, the difference between $\langle |Q^T f(n)| \rangle_p < \infty$ and Condition (B1) in the context of Theorem 1 is only whether we require p to have d or $d + 1$ rational moments.

Appendix C: Proofs of Section IV B

The proof of Theorem 2 builds on the following technical lemma.

Lemma C.1. *Suppose that \mathcal{P}_c^d is non-empty. Then, \mathcal{B}^r is non-empty for each $r = 0, 1, \dots$. Let $z^r \in \mathcal{B}^r$ for $r = 0, 1, \dots$. Identify the vectors z^r with vectors in \mathbb{R}^∞ by padding them with zeros. For each $A \subseteq \mathbb{N}^S$, the sequence $\{1_A^T z^r\}_{r \geq 0}$ is relatively compact. Furthermore, for each convergent subsequence $\{1_A^T z^{r_j}\}_{j \geq 0}$ we have that*

$$\lim_{j \rightarrow \infty} 1_A^T z^{r_j} = 1_A^T p$$

where $p \in \mathcal{P}_c^d$. The limiting distribution p depends on the subsequence in question, but not on the set A .

Proof. Pick any $p \in \mathcal{P}_c^d$. From (24) and the definition of \mathcal{P}_c^d it follows that $\tilde{P}_r p \in \mathcal{B}^r$ and consequently \mathcal{B}^r is non-empty. Now, let z^r be as in the premise and metrize \mathbb{N}^S using the discrete metric (which makes any function $f : \mathbb{N}^S \rightarrow \mathbb{R}$ continuous). Pick an r_* such that $c/r_*^d \leq 1/2$ and define

$$p^r := \frac{z^r}{1^T z^r}, \quad \forall r = r_*, r_* + 1, \dots$$

Note that p^r is a probability distribution. Additionally,

$$\langle |n|_w^d \rangle_{p^r} = \frac{1}{1^T z^r} \sum_{n \in \mathbb{N}_r} |n|_w^d z_n^r \leq 2c.$$

Since $n \mapsto |n|_w^d$ is an inf-compact function on \mathbb{N}^S , Proposition 1.4.15 in Ref.³¹, combined with the above uniform bound, shows that the sequence $(p^r)_{r \geq r_*}$ is tight. By Prohorov's Theorem, Theorem 1.4.12 in Ref.³¹, we have that $(p^r)_{r \geq r_*}$ is relatively compact in the sense that every sub-sequence of $(p^r)_{r \geq r_*}$ has a subsequence that converges weakly to a probability distribution on \mathbb{N}^S . Let p be any of these accumulation points and $(p^{r_k})_{k \geq 0}$ be a subsequence that converges to p . By weak convergence we have that for each n

$$(Qp)_n = \lim_{k \rightarrow \infty} (Q_{r_k} p^{r_k})_n = \lim_{k \rightarrow \infty} \frac{1}{1^T z^{r_k}} (Q_{r_k} z^{r_k})_n = 0.$$

Since, non-emptiness of \mathcal{P}_c^d implies that underlying Markov chain is regular (at least when restricted to the set of initial conditions relevant to the stationary distributions)¹², the fact that $Qp = 0$, implies that p is a stationary distribution of the network, see Theorem 4.3 and the discussion after Theorem 4.5 in Chapter 5.4

of Ref.³⁸. Additionally, since $n \mapsto |n|_w^d$ is a continuous function on \mathbb{N}^S that is bounded from below, Proposition 1.4.18 in Ref.³¹ shows that

$$\left\langle |n|_w^d \right\rangle_p \leq \liminf_{k \rightarrow \infty} \left\langle |n|_w^d \right\rangle_{p^{r_k}}.$$

Additionally,

$$\left\langle |n|_w^d \right\rangle_{p^{r_k}} = \frac{1}{1^T z^{r_k}} \sum_{n \in \mathcal{N}^{r_k}} |n|_w^d z_n^{r_k} \leq \frac{c}{1 - c/r_k^d}$$

Thus $\left\langle |n|_w^d \right\rangle_p \leq c$ and we have that $p \in \mathcal{P}_c^d$.

We now argue that $\{1_A^T z^r\}_{r \geq 0}$ is relatively compact and that it has the desired accumulation points. Pick any subsequence $\{1_A^T z^{r_j}\}_{j \geq 0}$ of $\{1_A^T z^r\}_{r \geq 0}$, and let $\{p^r\}_{r \geq 0}$ be the associated subsequence of $\{p^r\}_{r \geq 0}$. Due to relative compactness of $\{p^r\}_{r \geq 0}$, we can find a convergent subsequence $\{p^{r_{j_k}}\}_{k \geq 0}$ of $\{p^{r_j}\}_{j \geq 0}$ and let $p \in \mathcal{P}_c^d$ be its limit. Weak convergence of $\{p^{r_{j_k}}\}_{k \geq 0}$ and the fact that $1 - c/r^d \leq 1^T z^r \leq 1$ implies that $\{1_A^T z^{r_{j_k}}\}_{k \geq 0}$ converges too and that

$$\lim_{k \rightarrow \infty} 1_A^T z^{r_{j_k}} = \lim_{k \rightarrow \infty} (1^T z^{r_{j_k}})(1_A^T p^{r_{j_k}}) = 1_A^T p.$$

Note that p in the above depends on the subsequence $\{p^{r_{j_k}}\}_{k \geq 0}$ but not on A . Since the above holds for any subsequence of $\{1_A^T z^r\}_{r \geq 0}$, we have argued that $\{1_A^T z^r\}_{r \geq 0}$ is relatively compact and that it has the desired accumulation points. \square

We are now in a position to prove Theorem 2.

Proof of Theorem 2. 1. As pointed out in the proof of Lemma C.1, it is the case that if $p \in \mathcal{P}_c^d$, then $\tilde{P}_r p \in \mathcal{B}^r$ for each $r \geq 0$, and the result follows.

2. **The lower bounds.** Because the set \mathcal{B}^r is compact (from its definition it is closed, and it is contained in the hypercube $[0, 1]^{\#_r^w}$), and because $z \mapsto 1_A^T z$ is a linear function on $\mathbb{R}^{\#_r^w}$, for each $r \geq 0$, there exists a $z^r \in \mathcal{B}^r$ that achieves the infimum; $1_A^T z^r = l_A^r$. Due to 1. we have that

$$1_A^T z^r \leq l_A. \quad (\text{C1})$$

Lemma C.1, tells us that every accumulation point $\{1_A^T z^r\}_{r \geq 0}$ is of the form $1_A^T p$ where $p \in \mathcal{P}_c^d$ (the p may depend on the subsequence in question). Thus we have that

$$\liminf_{r \rightarrow \infty} 1_A^T z^r \geq \inf\{1_A^T p : p \in \mathcal{P}_c^d\} = l_A.$$

Combining the above with Eq. (C1) tells us that $\{1_A^T z^r\}_{r \geq 0}$ converges and that its limit is l_A .

The upper bounds: The proof is analogous to that of the lower bounds, one just needs to use

$$1_A^T z^r + \frac{c}{r^d} \geq u_A,$$

instead of Eq. (C1). \square

The following is a simple corollary of Theorem C with useful consequences that we briefly discuss in Section VI

Corollary C.2. *As in Lemma C.1, in this corollary we are identifying vectors $z \in \mathcal{B}^r$ with vectors in \mathbb{R}^∞ by padding them with zeros.*

1. Suppose that \mathcal{P}_c^d is the singleton $\{p\}$ and let $z^r \in \mathcal{B}^r$. Then, the sequence $\{z^r\}_{r \geq 0}$ converges pointwise to p . That is, for each $n \in \mathbb{N}^S$, the sequence $\{z_n^r\}_{r \geq 0}$ converges to p_n .
2. Suppose that there exists a unique $p \in \mathcal{P}_c^d$ such that $l_A = 1_A^T p$ (resp. $u_A = 1_A^T p$) and let $z^r \in \mathcal{B}^r$ be such $1_A^T z^r = l_A^r$ (resp. $1_A^T z^r = u_A^r$). Then, the sequence $\{z^r\}_{r \geq 0}$ converges pointwise to p .

Proof. 1. We prove this statement by contradiction. Uniqueness of p and Lemma C.1 tells us that, for each n , $\{z_n^r\}_{r \geq 0}$ is relatively compact and that p_n is the limit point of any converging subsequence. Suppose that $\{z_n^r\}_{r \geq 0}$ does not converge to p_n . This means that we can find an $\varepsilon > 0$ and a subsequence $\{z_n^{r_j}\}_{j \geq 0}$ such that $|z_n^{r_j} - p_n| > \varepsilon$ for all j . But, by relative compactness of $\{z_n^r\}_{r \geq 0}$, the subsequence $\{z_n^{r_j}\}_{j \geq 0}$ has a convergent subsequence and its limit must be p_n , giving us a contradiction.

2. We prove the statement for the lower bounds, the proof for the upper bounds is analogous. Theorem 2 tells us that $\{1_A^T z^r\}_{r \geq 0}$ converges to $1_A^T p$. Lemma C.1 then tells us that, for each $n \in \mathbb{N}^S$, $\{z_n^r\}_{r \geq 0}$ is convergent too with limit p_n . \square

The proofs of Lemmas 3 and 4 are very similar:

Proof of Lemma 3. Using the fact that $u^r \geq P_r p$ componentwise and Eq. (24) we have that

$$\|u^r - \tilde{P}_r p\|_1 = 1^T (u^r - \tilde{P}_r p) = 1^T u^r - 1 + \mu_r \leq 1^T u^r - 1 + \frac{c}{r^d}.$$

Similarly, using the fact that $l^r \leq \tilde{P}_r p$ componentwise we have that

$$\|\tilde{P}_r p - l^r\|_1 = 1^T (\tilde{P}_r p - l^r) = 1 - 1^T l^r - \mu_r \leq 1 - 1^T l^r. \quad \square$$

Proof of Lemma 4. Let

$$\tilde{\mathcal{N}}^r : \{n \in \mathbb{N}^{\tilde{S}} : \exists m \in \mathbb{N}^{\tilde{S}}, (n, m) \in \mathcal{N}^R\},$$

be the projection of \mathcal{N}^r onto $\mathbb{N}^{\tilde{S}}$. By Theorem 2 1. we have that $\sum_{m \in \mathcal{N}_n} p(n, m) \leq \tilde{u}_n^r$. Consequently

$$\begin{aligned} \|\tilde{u}^r - \tilde{P}_r \tilde{p}\|_1 &= \sum_{n \in \tilde{\mathcal{N}}^r} |\tilde{u}_n^r - \tilde{p}_n| = \sum_{n \in \tilde{\mathcal{N}}^r} \left| \tilde{u}_n^r - \sum_{m \in \mathbb{N}^{\tilde{S}}} p(n, m) \right| \\ &= \sum_{n \in \tilde{\mathcal{N}}^r} \left(\tilde{u}_n^r - \sum_{m \in \mathcal{N}_{r,n}} p(n, m) \right) + \sum_{n \in \tilde{\mathcal{N}}^r} \sum_{m \in (\mathcal{N}_{r,n})^c} p(n, m) \\ &\leq 1^T \tilde{u}^r - 1 + 2\mu_r \leq 1^T \tilde{u}^r - 1 + \frac{2c}{r^d}, \end{aligned}$$

where the first inequality follows from the facts that

$$\bigcup_{n \in \mathcal{N}^r} \bigcup_{m \in \mathcal{N}_{r,n}} \{(n, m)\} = \mathcal{N}_r,$$

$$\bigcup_{n \in \mathcal{N}^r} \bigcup_{m \in \mathcal{N}_{r,n}^c} \{(n, m)\} \subseteq (\mathcal{N}_r)^c,$$

and the second from Eq. (24).

Since $\tilde{p}_n \geq \tilde{l}_n^r$ and $\tilde{p}_n \leq \tilde{u}_n^{*r}$, the proofs of Eqs. (27a) and (27c) are identical to those of Eqs. (25a) and (25b). \square

- ¹M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, “Stochastic gene expression in a single cell,” *Science* **297**, 1183–1186 (2002).
- ²Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, “Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells,” *Science* **329**, 533–538 (2010).
- ³S. Uphoff, N. D. Lord, B. Okumus, L. Potvin-Trottier, D. J. Sherratt, and J. Paulsson, “Stochastic activation of a DNA damage response causes cell-to-cell mutation rate variation,” *Science* **351**, 1094–1097 (2016).
- ⁴A. Arkin, J. Ross, and H. H. McAdams, “Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected *Escherichia coli* cells,” *Genetics* **149**, 1633–1648 (1998).
- ⁵S. H. Dandach and M. Khammash, “Analysis of stochastic strategies in bacterial competence: a master equation approach,” *PLoS Comput Biol* **6**, e1000985 (2010).
- ⁶P. Thomas, N. Popović, and R. Grima, “Phenotypic switching in gene regulatory networks,” *Proc Natl Acad Sci* **111**, 6994–6999 (2014).
- ⁷B. Munsky, Z. Fox, and G. Neuert, “Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics,” *Methods* **85**, 12–21 (2015).
- ⁸F. Fröhlich, P. Thomas, A. Kazeroonian, F. J. Theis, R. Grima, and J. Hasenauer, “Inference for stochastic chemical kinetics using moment equations and system size expansion,” *PLoS Comput Biol* **12**, e1005030 (2016).
- ⁹G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden, “Systematic identification of signal-activated stochastic gene regulation,” *Science* **339**, 584–587 (2013).
- ¹⁰D. A. Oyarzun, J.-B. Lugagne, and G.-B. V. Stan, “Noise propagation in synthetic gene circuits for metabolic control,” *ACS Synth Biol* **4**, 116–125 (2014).
- ¹¹C. Zechner, G. Seelig, M. Rullan, and M. Khammash, “Molecular circuits for dynamic noise filtering,” *Proc Natl Acad Sci* **113**, 4729–4734 (2016).
- ¹²S. P. Meyn and R. L. Tweedie, “Stability of Markovian processes II: Continuous-time processes and sampled chains,” *Adv Appl Probab* **25**, 487–517 (1993).
- ¹³M. Hemberg and M. Barahona, “Perfect sampling of the master equation for gene regulatory networks,” *Biophys J* **93**, 401–410 (2007).
- ¹⁴C. Gardiner, *Stochastic Methods*, 4th ed. (Springer, Berlin, 2009).
- ¹⁵T. Jahnke and W. Huisinga, “Solving the chemical master equation for monomolecular reaction systems analytically,” *J Math Biol* **54**, 1–26 (2007).
- ¹⁶F. López-Caamal and M. Tatiana, “Exact probability distributions of selected species in stochastic chemical reaction networks,” *Bull Math Biol* **76**, 2334–2361 (2014).
- ¹⁷N. Van Kampen, “The expansion of the master equation,” *Adv Chem Phys* **34**, 245–309 (1976).
- ¹⁸V. Danos and N. Oury, “Equilibrium and termination II: the case of Petri Nets,” *Math Structures Comput Sci* **23**, 290–307 (2013).
- ¹⁹D. F. Anderson, G. Craciun, and T. G. Kurtz, “Product-form stationary distributions for deficiency zero chemical reaction networks,” *Bull Math Biol* **72**, 1947–1970 (2010).
- ²⁰D. Cappelletti and C. Wiuf, “Product-form Poisson-like distributions and complex balanced reaction systems,” *SIAM J Appl Math* **76**, 411–432 (2016).
- ²¹J. Peccoud and B. Ycart, “Markovian modeling of gene-product synthesis,” *Theor Popul Biol* **48**, 222–234 (1995).
- ²²P. Bokes, J. R. King, A. T. Wood, and M. Loose, “Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression,” *J Math Biol* **64**, 829–854 (2012).
- ²³N. Van Kampen, “The equilibrium distribution of a chemical mixture,” *Phys Lett A* **59**, 333–334 (1976).
- ²⁴S. Engblom, “Computing the moments of high dimensional solutions of the master equation,” *Appl Math Comput* **180**, 498–515 (2006).
- ²⁵E. Lakatos, A. Ale, P. D. Kirk, and M. P. Stumpf, “Multivariate moment closure techniques for stochastic kinetic models,” *J Chem Phys* **143**, 094107 (2015).
- ²⁶D. Schnoerr, G. Sanguinetti, and R. Grima, “Comparison of different moment-closure approximations for stochastic chemical kinetics,” *J Chem Phys* **143**, 185101 (2015).
- ²⁷S. Engblom and J. Pender, “Approximations for the moments of nonstationary and state dependent birth-death queues,” *arXiv preprint arXiv:1406.6164* (2014).
- ²⁸T. G. Kurtz, “Limit theorems and diffusion approximations for density dependent Markov chains,” in *Stochastic Systems: Modeling, Identification and Optimization, I* (Springer, 1976) pp. 67–78.
- ²⁹E. Schwerer, *A linear programming approach to the steady-state analysis of Markov processes*, Ph.D. thesis, Graduate School of Business, Stanford University (1996).
- ³⁰K. Helmes and R. H. Stockbridge, “Extension of Dale’s moment conditions with application to the Wright–Fisher model,” *Stoch Model* **19**, 255–267 (2003).
- ³¹O. Hernández-Lerma and J. B. Lasserre, *Markov Chains and Invariant Probabilities* (Birkhäuser, Basel, 2003).
- ³²J. Kuntz, M. Ottobre, G.-B. Stan, and M. Barahona, “Bounding stationary averages of polynomial diffusions via semidefinite programming,” *SIAM J Sci Comput* **38**, A3891–A3920 (2016).
- ³³S. Engblom, “Spectral approximation of solutions to the chemical master equation,” *J Comput Appl Math* **229**, 208–221 (2009).
- ³⁴P. Thomas and R. Grima, “Approximate probability distributions of the master equation,” *Phys Rev E* **92**, 012120 (2015).
- ³⁵B. Munsky and M. Khammash, “The finite state projection algorithm for the solution of the chemical master equation,” *J Chem Phys* **124**, 044104 (2006).
- ³⁶V. Kazeev, M. Khammash, M. Nip, and C. Schwab, “Direct solution of the chemical master equation using quantized tensor trains,” *PLoS Comput Biol* **10**, e1003359 (2014).
- ³⁷D. T. Gillespie, “Exact stochastic simulation of coupled chemical reactions,” *J Phys Chem* **81**, 2340–2361 (1977).
- ³⁸W. J. Anderson, *Continuous-time Markov chains: An applications-oriented approach*, 1st ed. (Springer, New York, 1991).
- ³⁹S. P. Meyn and R. L. Tweedie, “Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes,” *Adv Appl Probab* , 518–548.
- ⁴⁰A. Gupta and M. Khammash, “A scalable method for finding irreducible state-spaces for stochastic models of biochemical reaction networks,” *ArXiv:1505.06594* (2015).
- ⁴¹F. Schlögl, “Chemical reaction models for non-equilibrium phase transitions,” *Z Phys* **253**, 147–161 (1972).
- ⁴²W. Ebeling and L. Schimansky-Geier, “Stochastic dynamics of a bistable reaction system,” *Physica A* **98**, 587–600 (1979).
- ⁴³J. B. Lasserre, *Moments, Positive Polynomials and Their Applications*, Imperial College Press Optimization Series (Imperial College Press, London, 2009).
- ⁴⁴G. Blekherman, P. A. Parrilo, and R. R. Thomas, *Semidefinite*

- Optimization and Convex Algebraic Geometry* (SIAM, Philadelphia, 2013).
- ⁴⁵J. Löfberg, “Yalmip: A toolbox for modeling and optimization in matlab,” in *Proc IEEE Int Symp Comput Aid Control Syst Des* (IEEE, 2004) pp. 284–289.
- ⁴⁶M. Nakata, “A numerical evaluation of highly accurate multiple-precision arithmetic version of semidefinite programming solver: SDPA-GMP, -QD and -DD,” in *Proc IEEE Int Symp Comput Aid Control Syst Des* (IEEE, 2010) pp. 29–34.
- ⁴⁷G. Fantuzzi, *mpYALMIP (Version 1.0)* (2016), available at <https://github.com/giofantuzzi/mpYALMIP>.
- ⁴⁸G. Fantuzzi, D. Goluskin, D. Huang, and S. I. Chernyshenko, “Bounds for deterministic and stochastic dynamical systems using sum-of-squares optimization,” *SIAM J Appl Dyn Syst* **15**, 1962–1988 (2016).
- ⁴⁹L. Paulevé, G. Craciun, and H. Koepl, “Dynamical properties of discrete reaction networks,” *J Math Biol* **69**, 55–72 (2014).
- ⁵⁰T. S. Gardner, C. R. Cantor, and J. J. Collins, “Construction of a genetic toggle switch in *Escherichia coli*,” *Nature* **403**, 339–342 (2000).
- ⁵¹R. Perez-Carrasco, P. Guerrero, J. Briscoe, and K. M. Page, “Intrinsic noise profoundly alters the dynamics and steady state of morphogen-controlled bistable genetic switches,” *PLoS Comput Biol* **12**, e1005154 (2016).
- ⁵²CPLEX, IBM ILOG, *V12.1: Users Manual for CPLEX* (2009), International Business Machines Corporation.
- ⁵³T. Dayar, H. Hermanns, D. Spieler, and V. Wolf, “Bounding the equilibrium distribution of Markov population models,” *Numer Linear Algebra Appl* **18**, 931–946 (2011).
- ⁵⁴D. Spieler, *Numerical analysis of long-run properties for Markov population models*, Ph.D. thesis, Saarland University (2014).
- ⁵⁵A. Miliadis-Argeitis and M. Khammash, “Optimization-based Lyapunov function construction for continuous-time Markov chains with affine transition rates,” in *Proc IEEE Conf Decis Control* (IEEE, 2014).
- ⁵⁶P. W. Glynn and A. Zeevi, “Bounding stationary expectations of Markov processes,” in *Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz*, Vol. 4, edited by E. N. Stewart, J. Feng, and R. H. Stockbridge (Institute of Mathematical Statistics, 2008) pp. 195–214.
- ⁵⁷Z. Fox, G. Neuert, and B. Munsky, “Finite state projection based bounds to compare chemical master equation models using single-cell data,” *J Chem Phys* **145**, 074101 (2016).